5K

УДК 004.89:004.93

Ю.Г. Кривонос, Ю.В. Крак, А.В. Бармак, А.С. Загваздин
Институт кибернетики им. В.М. Глушкова НАН Украины, г. Киев, Украина
yuriy.krak@gmail.com

Автоматизированная система стенографирования

В статье рассматривается проблема построения автоматизированной системы стенографирования, предназначенной для оптимизации работы операторов-стенографистов в организациях и учреждениях, где требуется ведение текстовых стенограмм заседаний, собраний, докладов, конференций. Авторами рассмотрены преимущества и недостатки существующих программных продуктов и систем и предложен прототип системы автоматизированного стенографирования.

Введение

Текстовая стенограмма заседания является необходимой составляющей в работе многих организаций. Как правило, процесс создания и расшифровки стенограмм достаточно продолжителен и попытки его ускорения путем расширения персонала, вовлеченного в процесс, представляются неэффективными. Для автоматизации процесса создания стенограмм заседаний предлагается рассматриваемая система распределенного компьютерного документирования.

Поскольку стенографирование заседаний является задачей, которая может быть достаточно просто распределена между многими исполнителями, система распределенного стенографирования должна поддерживать однопользовательский и многопользовательский режимы работы. На сегодняшний день в мире существует несколько систем распределенного документирования, однако практически всем им свойственнен ряд недостатков, от которых авторы стремились избавить предложенную систему. Основным недостатком существующих систем стенографирования заседаний является то, что они предъявляют высокие требования к аппаратному обеспечению, на котором работает серверная часть системы и существенные затраты на внедрение системы в целом. Такие системы зачастую требуют постоянного системного администрирования [1-3], что ограничивает внедрение подобных систем стенографирования в небольших организациях, организациях с ограниченным ИТ бюджетом, а также для индивидуальных пользователей. Наличие серверной части также делает практически невозможной мобильную работу с системой и требует постоянного подключения к локальной сети.

Предложенная в статье система распределенного стенографирования предназначена автоматизировать и упростить работу индивидуальных операторов-стенографистов и групп стенографистов в организациях различного уровня и является продолжением работы над системой автоматизированного стенографирования [4]. Очевидным является факт, что навыки работы с компьютером операторов-стенографистов в большинстве организаций являются относительно ограниченными, что накладывает особые требования на пользовательский интерфейс и эргономику системы автоматизированного стенографирования. В частности, как можно большая часть операций в такой системе

должна быть автоматизирована и не требовать каких-либо действий от пользователя, а набор операций, где действие пользователя необходимо, должен быть ограниченным и состоять из небольшого числа простых и понятных операций. При этом система должна оставаться гибкой и поддерживать достаточно большое количество настроек, чтобы сделать работу оператора как можно более простой.

Требования к системе распределенного стенографирования и характеристики предложенной системы

Эффективная система автоматизированного стенографирования должна удовлетворять следующим требованиям:

– Получать и сохранять звуковой сигнал с речью, предназначенной для стенографирования, поддерживать значительную часть современных форматов хранения звуковой информации.
– Реализовывать предварительную цифровую обработку звукового сигнала, в частности обесшумливание сигнала.
– Выполнять автоматическую разбивку сигнала на сегменты, продолжительность которых была бы максимально удобной для обработки стенографистом.
– Поддерживать однопользовательский и многопользовательский режимы работы системы, при этом не требуя комплексного внедрения и сопровождения в рамках организации.
– В многопользовательском режиме эффективно распределять сегменты речевого сигнала между операторами-стенографистами.

Очевидным является факт о том, что уровень компьютерной грамотности операторов-стенографистов традиционно не высок, что предъявляет особые требования к эргономике системы и удобству пользовательского интерфейса.

Рассматриваемая система обладает следующими характеристиками:

– Пользовательский интерфейс соответствует основным требованиям, которые традиционно предъявляются к эргономике и интерфейсу программного продукта. В частности, согласно классическому исследованию Д. Миллера про кратковременную память человека, который утверждает, что человек может одновременно концентрировать внимание лишь на 7+/-2 объектах, количество основных команд, доступных пользователю при работе с речевым сигналом, сведено к 8, а оптимальная продолжительность сегмента звукового сигнала определена на уровне 5 – 9 слов. К тому же для всех основных команд в интерфейсе реализованы «горячие клавиши», с целью дать пользователю возможность управлять программным продуктом, не отрывая рук от клавиатуры и не прекращая набора текста стенограммы.
– Система выполняет эффективную разбивку звукового сигнала на эквивалентные сегменты. Такая разбивка реализовывается согласно следующему принципу: алгоритм разбивки проходит по сигналу окном заданной длины и определяет фрагменты сигнала, соответствующие паузам в речевом потоке (среднеквадратичное отклонение в таких сегментах не превышает заданной пороговой величины). Согласно найденным паузам определяются границы сегмента с условием, что продолжительность сегмента должна быть не меньше заданной величины и чтобы количество пауз в рамках сегмента соответствовало оптимальному количеству слов в сегменте. Паузы, продолжительность которых превышает заданный порог, вырезаются из рассматриваемого сигнала.

– Система работает как в однопользовательском, так и в многопользовательском режимах.

– Среди объединенных в сеть компьютеров, на которых установлена система, один определяется как главный. Он выполняет роль сервера, обеспечивает получение сигнала, его предварительную цифровую обработку, разбивку сигнала на сегменты и распределение сегментов между остальными компьютерами. Таким образом система практически не требует внедрения и администрирования.

– При предварительной обработке звукового сигнала система изменяет амплитуду звукового сигнала, чтобы уровень громкости оставался комфортным для стенографирования, изменяет частоту основного тона и скорость воспроизведения сигнала.

– В многопользовательском режиме система эффективно распределяет сегменты между операторами-стенографистами. На главном компьютере сегменты, полученные в результате предварительной обработки, организованы в виде очереди, при освобождении оператора система направляет ему следующий в очереди сегмент, нуждающийся в расшифровке.

Интерфейс АРМ оператора-стенографиста представлен на рис. 1.

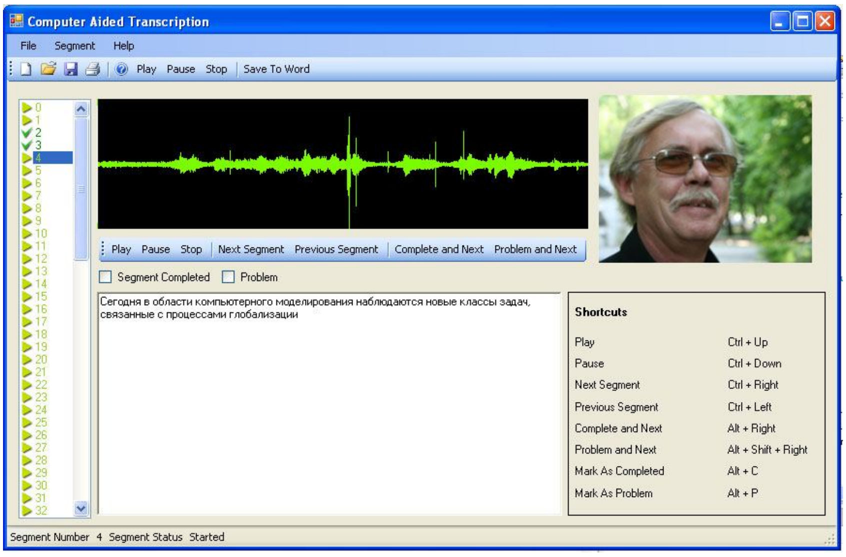


Рисунок 1 – Интерфейс АРМ оператора-стенографиста

Предварительная цифровая обработка сигнала

Важным аспектом для системы автоматизированного стенографирования является возможность изменения скорости воспроизведения звукового сигнала, возможность воспроизведения звукового сигнала быстрее или медленнее без утраты его основных акустических характеристик: частоты основного тона и амплитуды. Для манипуляций над частотой основного тона и продолжительностью сигнала для создания эффектов медленного или быстрого воспроизведения в данной работе предлагается использование PSOLA-подобных алгоритмов. Для реализации таких алгоритмов сперва решается задача обнаружения периодов псевдопериодичности в звуковом сигнале (рис. 2). Для этого исходный звуковой сигнал пропускается через низкочастотный и высокочастотный фильтры с конечными импульсными характеристиками. Результат применения такого рода фильтров к звуковому сигналу, соответствующему слогу «ма», приведен на рис. 3.

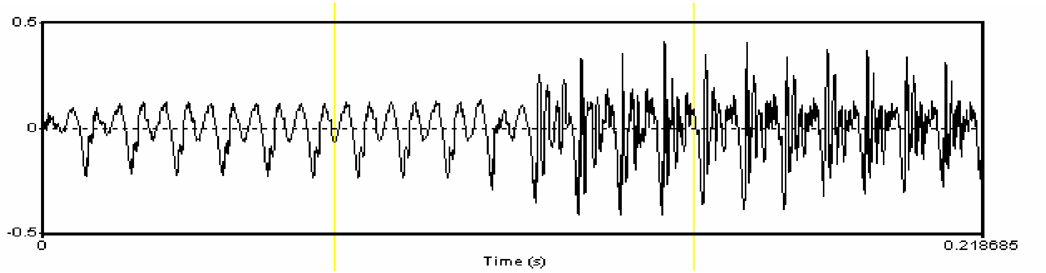


Рисунок 2 – Слог «ма» до применения фильтров

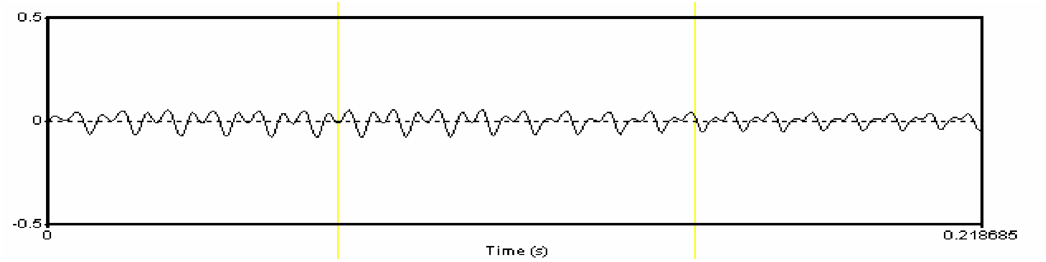


Рисунок 3 – Слог «ма» после фильтрации

Далее для сглаживания сигнала каждый элемент вектора исходного сигнала заменяется на взвешенное среднее четырех окружающих его элементов по формуле:

$$d[i] = \frac{3x[i-2] + x[i-1] - x[i+1] - 3x[i+2]}{10}. \quad (1)$$

К полученному сигналу применяется медианное сглаживание порядка $n = 199$, (каждый элемент вектора заменяется на медиану вектора, состоящего из n элементов, окружающих текущий элемент). Вид сигнала после сглаживания представлен на рис. 4.

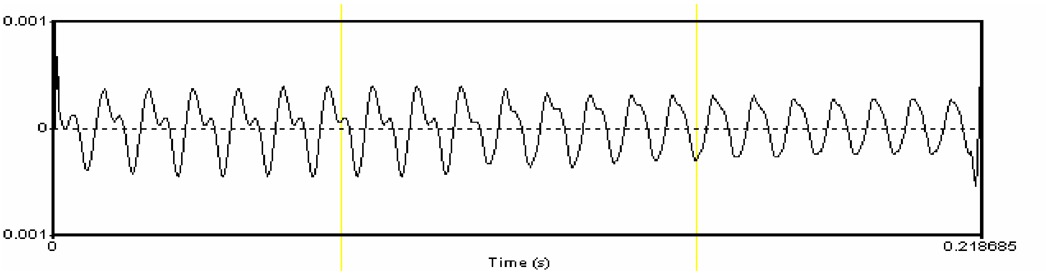


Рисунок 4 – Слог «ма» после фильтрации и медианного сглаживания

После этого в полученном сигнале обнаруживаются точки, где последовательность, состоящая из элементов вектора сигнала, изменяет знак с – на + и такие точки обозначаются как границы питч-периодов. Среди определенных таким образом границ обнаруживаются и исключаются точки, расположенные слишком близко друг к другу, а для участков сигнала, где нет явной псевдопериодичности, назначаются условные границы с некоторым постоянным интервалом.

После определения границ периодов псевдопериодичности можно изменять акустические характеристики сигнала. Исходный сигнал можно представить в виде функции периодов основного тона $x_i[n]$:

$$x[n] = \sum_{i=-\infty}^{\infty} x_i[n - t_a[i]], \quad (2)$$

где $t_a[i]$ – границы периодов псевдопериодичности сигнала, т.е. разница между двумя соседними границами $P_a[i] = t_a[i] - t_a[i-1]$ равняется периоду основного тона в момент времени $t_a[i]$. Питч-период определим через исходный сигнал, помноженный на оконную функцию:

$$x_i[n] = w_i[n]x[n], \tag{3}$$

где окна w_i удовлетворяют условию:

$$\sum_{i=-\infty}^{\infty} w_i[n - t_a[i]] = 1, \tag{4}$$

что достигается использованием оконных функций типа Хэннинга или трапециевидным окном длиной в два периода основного тона.

В результате работы алгоритма необходимо получить сигнал $y[n]$, который имеет одинаковые с $x[n]$ спектральные характеристики, но отличается от него основным тоном и/или продолжительностью. Чтобы достичь этого, заменяем аналитические границы питч-периодов $t_a[i]$ границами $t_b[i]$, а аналитические периоды основного тона $x_i[n]$ периодами $y_i[n]$ согласно

$$y[n] = \sum_{j=-\infty}^{\infty} y_j[n - t_b[j]]. \tag{5}$$

Таким образом, теперь достаточно лишь задать границы $t_b[i]$, которые соответствуют продолжительности и основному тону, которые необходимо получить. Результирующий период основного $y_i[n]$ тона получаем подстановкой ближайшего соответствующего аналитического периода $x_i[n]$. Графически работа алгоритма представлена на рис. 5.

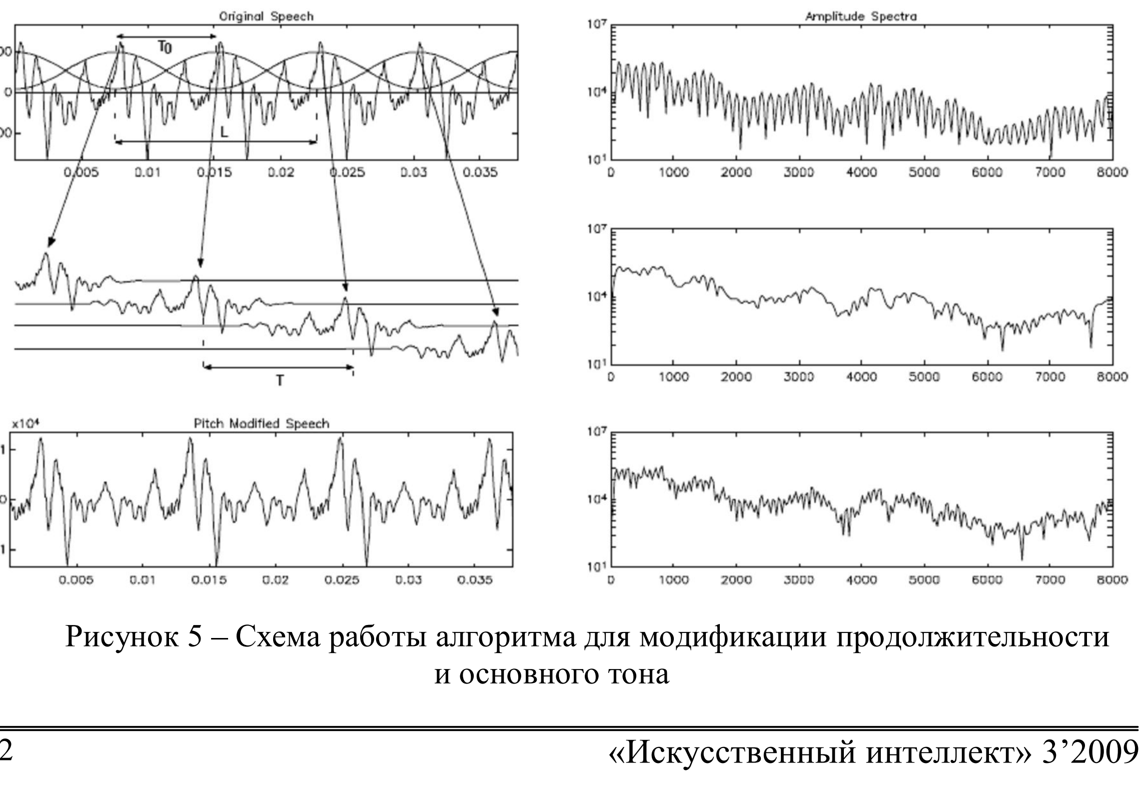


Рисунок 5 – Схема работы алгоритма для модификации продолжительности и основного тона

Выводы

После реализации прототипа системы был проведен эксперимент, целью которого было сравнение эффективности работы одного стенографиста и группы стенографистов при использовании предложенной системы и без такового, то есть при использовании лишь традиционных средств (таких как Windows Media Player для воспроизведения звукового сигнала и Microsoft Office Word для набора текста стенограммы). В качестве входящего речевого сигнала для эксперимента была выбрана запись заседания специализированного ученого совета по поводу защиты диссертации продолжительностью около 2 часов. В результате эксперимента получены следующие результаты:

– При работе одного оператора-стенографиста на стенографирование записи с использованием предложенной системы было потрачено около 4 часов. Для стенографирования этой же записи при использовании стандартных средств оператор тратит в среднем 12 – 16 часов.

– Группе стенографистов из 5 человек для обработки записи понадобилось около 40 минут, после чего полученный текст стенограммы был направлен на обработку корректору. Вместе с коррекцией в общей сложности расшифровка стенограммы заняла около 1 часа.

Проведенный эксперимент демонстрирует эффективность предложенной системы в сравнении с традиционными средствами. Вместе с остальными преимуществами, среди которых отсутствие необходимости администрирования и внедрения, простота в использовании и качественная предварительная обработка сигнала, предложенная система является достаточно эффективным средством для автоматизации процесса создания и расшифровки стенограмм заседаний для небольших и крупных организаций, а также для индивидуальных пользователей.

Литература

1. Система распределенного компьютерного документирования устных выступлений и фонограмм речи Нестор [Электронний ресурс]. – Режим доступа : http://www.speechpro.ru/rus/products/doc-systems/nestor/
2. Комплекс оперативного стенографирования «SRS Report 2002» [Электронний ресурс]. – Режим доступа : http://srs.kiev.ua/ index.php?pg=2&lang=rus&tov=23
3. The meeting recorder project – [Электронний ресурс]. – Режим доступа : http://www.icsi.berkeley.edu/Speech/mr/mtgrcdr.html
4. Кривонос Ю.Г. Система автоматизації отримання стенограм засідань / Ю.Г. Кривонос, Ю.В. Крак, О.В. Бармак, О.С. Загваздін // Штучний інтелект. – 2008.

Ю.Г. Кривонос, Ю.В. Крак, О.В. Бармак, О.С. Загваздін

Автоматизована система стенографування

У статті розглядається проблема побудови автоматизованої системи стенографування, що призначена для оптимізації роботи операторів-стенографістів в організаціях і установах, де необхідне ведення текстових стенограм засідань, доповідей, конференцій. Авторами розглядаються переваги і недоліки існуючих систем і пропонується прототип системи автоматизованого стенографування.

Yu.G. Kryvonos, Yu.V. Krak, O.V. Barmak, O.S. Zagvazdin

Automated Transcription System

The paper is devoted to problem of automated transcription system construction intended to optimize the transcription processes involving transcription operators in organizations where transcription is required for meetings, reports, conferences. Authors analyze pros and cons of existing systems and suggest a prototype of automated transcription system.

Статья поступила в редакцию 09.06.2009.